MATH 279 Final Report: Learning Cross-Asset Order Flow Networks for Daily Return Prediction

Brendan Connelly[1] and Rigel Mummers[1]

[1]Department of Mathematics, University of California, Los Angeles, 90095, CA. Email(s): brendanconnelly@ucla.edu, rigelmummers@ucla.edu

Supervisor(s): Mihai Cucuringu
April 12, 2026

Introduction A large body of market microstructure research supports the use of order flow imbalance (OFI) as a predictor of short-term returns. Empirical studies show that imbalances between buy- and sell-side pressure convey information about latent supply–demand dynamics and are strongly associated with immediate price changes and short-horizon returns. In particular, Cont et al. (2013) shows that OFI explains a substantial portion of high-frequency price variation and often outperforms more traditional volume- or trade-based imbalance measures.

Our goal is to test whether intraday order flow contains information that persists beyond the trading day and can be used to predict either overnight close-to-open returns or next-day open-to-close returns across assets. We not only explore the self-predictive content of an asset's own OFI, but, perhaps more importantly, *cross-asset* effects: whether the order flow signal of one stock helps predict the return of another stock. This naturally leads to a matrix-based formulation, where we learn a cross-impact matrix $W \in \mathbb{R}^{N\times N}$ mapping lagged OFI signals to next-day returns. We thus primarily consider linear models in this paper. Our main finding is that imposing sector-block structure and low-rank denoising on this matrix produces a strong and robust predictive signal, achieving annualized sector-neutral Sharpe ratios above 1.5.

The rest of the report is organized as follows. Section 1 reviews and compares to the relevant literature. Section 2 describes the data construction, including the OFI definition and daily signal aggregation. Section 3 presents a preliminary sparse directed-edge baseline that motivates the move to structured estimation. Section 4 develops the main methodology: ridge estimation of the cross-impact matrix under various structural constraints, and low-rank denoising via Gavish–Donoho singular value thresholding. Section 4.9 presents the main results. We conclude our findings in section 5.

1 Literature Review

Our project sits at the intersection of market microstructure, asset cross-impact, and factor-based learning. We first review the OFI and cross-impact literature, then discuss the clustering and structured estimation methods that informed our approach.

1.1 OFI, Cross-Impact, and Market Microstructure

The natural starting point is the single-asset OFI literature. A central result of Cont et al. (2013) is that short-horizon price changes are closely related to order flow imbalance defined from changes in bid and ask

queues at the best quotes. OFI captures the net pressure of limit orders, market orders, and cancellations, and explains price movements more robustly than simpler trade-imbalance or volume-based measures. This provides the main motivation for using OFI rather than raw trade counts.

At the daily horizon, Chordia and Subrahmanyam (2004) studied daily order imbalances and stock returns and found that lagged imbalances contain predictive information for future returns. Although their OFI variable is trade-imbalance-based rather than limit-order-book-based, the conceptual question is the same: whether persistent order pressure contains information beyond the contemporaneous price move.

Our project goes beyond the single-asset setting by asking whether the OFI of one asset predicts the return of another. In earlier microstructure work, Hasbrouck (1991) modeled trades and quote revisions jointly in a vector autoregressive framework, finding that a trade's price impact is lagged, concave in size, and larger when crossing a wider spread. More recently, Benzaquen et al. (2017) showed that cross-impact is empirically detectable in equity markets: while off-diagonal effects are weaker than own-asset impact, they offer some explanatory power. This is relevant because it suggests that cross-asset order-flow effects are real but likely weak and noisy.

Most directly relevant, Cont et al. (2023) studied cross-impact of OFI in a multi-asset setting. They found that including off-diagonal cross-asset information does not improve contemporaneous explanatory power—the immediate price impact of a stock is dominated by its own OFI. However, they do find predictive power in off-diagonal information at a short-term horizon, which is the key fact our project exploits. This is somewhat in tension with Capponi and Cont (2020), which emphasizes that off-diagonal entries do not contribute additional contemporaneous explanatory power.

1.2 Structured Estimation and Clustering

Our preliminary baseline uses false-discovery-rate control in the sense of Benjamini and Hochberg (1995) to sparsify directed edges, which has some success in terms of next day prediction.

The move to structured matrix estimation is informed by a biclustering heuristic approach. Dhillon (2001) studied simultaneous clustering of rows and columns of a matrix by modeling the data as a bipartite graph and applying spectral methods, showing that singular vectors of the normalized matrix provide a relaxation of a hard discrete partitioning problem. This biclustering perspective is relevant because the directed score matrix naturally has a row/column (target/source) interpretation.

He et al. (2022) developed DIGRAC, a directed graph clustering method that treats directionality as the main signal and clusters nodes by maximizing flow imbalance between groups. Bennett et al. (2022) proposed MetaCluster, which aggregates noisy pairwise lead–lag information into a directed cluster-level meta-flow matrix and ranks clusters by overall leadingness. Cartea et al. (2023) used pairwise lead–lag scores to extract globally leading and lagging assets.

While we did not ultimately employ these last clustering methods directly, they informed our decision to focus on structured block-level estimation. The sector-block constraint we impose on W can be seen as a hard, exogenously defined version of the biclustering that these methods would discover endogenously. Our results suggest that GICS sector membership already provides a strong and effective grouping for cross-impact estimation, though data-driven clustering as we have talked about in class remains a natural direction for future work, particularly when we have more deeper levels of OFI data.

For the low-rank denoising step, we use the optimal hard singular value threshold of Gavish and Donoho (2014), which provides a cutoff under a spiked random matrix model. This is well-suited to our setting because the estimated W matrix is expected to have a small number of strong latent factors contaminated by estimation noise. It is also computationally efficient to run.

2 Data Construction and Initial Setup

2.1 Order Flow Imbalance

A simple proxy for order flow is the trade imbalance,

$$\mathrm{TIB}_t = \frac{B_t - S_t}{B_t + S_t},$$

where B_t and S_t denote buyer- and seller-initiated trading volume over day t. However, as Cont et al. (2013) shows, trade imbalance is a noisy measure of supply–demand pressure because it ignores changes in displayed liquidity arising from limit order submissions and cancellations.

From Cont et al. (2013), order flow imbalance is defined using order book dynamics. Let $P_{i,n}^{\ell,b}$ and $P_{i,n}^{\ell,a}$ denote the ℓ-th best bid and ask prices at order book update n, with $q_{i,n}^{\ell,b}$ and $q_{i,n}^{\ell,a}$ the corresponding queue sizes. The level-ℓ bid and ask order flows are

$$OF_{i,n}^{\ell,b} := \begin{cases} q_{i,n}^{\ell,b}, & P_{i,n}^{\ell,b} > P_{i,n-1}^{\ell,b}, \\ q_{i,n}^{\ell,b} - q_{i,n-1}^{\ell,b}, & P_{i,n}^{\ell,b} = P_{i,n-1}^{\ell,b}, \\ -q_{i,n-1}^{\ell,b}, & P_{i,n}^{\ell,b} < P_{i,n-1}^{\ell,b}, \end{cases}$$

and $OF_{i,n}^{\ell,a}$ is defined symmetrically for the ask side. The level-ℓ OFI accumulated during the interval $(t-h, t]$ is

$$\mathrm{OFI}_{i,t}^{\ell,h} := \sum_{n=N(t-h)+1}^{N(t)} \left(OF_{i,n}^{\ell,b} - OF_{i,n}^{\ell,a}\right),$$

computed at minutely frequency ($h = 1$ minute). Stacking across the 10 levels gives the multi-level OFI vector $\mathbf{ofi}_{i,t}^{(h)} := (\mathrm{OFI}_{i,t}^{1,h}, \ldots, \mathrm{OFI}_{i,t}^{10,h})^\top \in \mathbb{R}^{10}$.

Because buying and selling pressure across LOB levels is highly correlated, we apply PCA to the matrix of 10-level OFI vectors pooled across all assets, days, and minutes. The first principal component $\mathbf{w}_1 \in \mathbb{R}^{10}$ captures the dominant direction of cross-level co-movement. We use the *integrated order flow imbalance* (IOFI) as the projection onto this component:

$$\mathrm{IOFI}_{i,t}^{h} = \frac{\mathbf{w}_1^\top \mathbf{ofi}_{i,t}^{(h)}}{\|\mathbf{w}_1\|_1},$$

as defined in Cont et al. (2023). This yields a single scalar per asset per minute—390 observations per trading day—that isolates the primary market signal from the 10-level book while suppressing level-specific noise. The results yielded were much stronger when using IOFI, so much so that we do not include any results without it.

2.2 Rank Normalization

Because raw daily OFI values are not necessarily designed to be directly comparable across assets, we also consider a cross-sectional rank normalization in some of our strategies. For each day t, we rank all N assets by their daily OFI signal and replace each value by its percentile rank in that day's cross-section. The transformed signal therefore reflects where an asset sits relative to the rest of the universe on that day, rather than its raw OFI magnitude.

2.3 Daily Signal Aggregation

For each asset i and day t, we aggregate the 390 minutely IOFI observations into a single daily signal. Letting k index minutes backward from the market close, we compute

$$\mathrm{ofi}_{i,t}^{\mathrm{daily}} = \sum_{k=1}^{390} e^{-\lambda k}\, \mathrm{IOFI}_{i,t,k}^{h},$$

where $\lambda > 0$ controls the exponential decay rate. Since we index backward from the close, larger k corresponds to earlier in the trading day, so this weighting puts more weight on order flow near the close and less weight on order flow further back in the day.

It is often more convenient and intuitive to write the decay in terms of a half-life τ, where

$$\lambda = \frac{\log 2}{\tau}.$$

In the implementation, we vary the half-life across a grid and treat it as a tuning parameter rather than fixing it in advance. We also considered a data-driven alternative: applying PCA to the 390-dimensional vector of minutely IOFI observations pooled across all assets and days, and using the first principal component as the aggregation kernel. The resulting weight profile resembles a kernel concentrated on the center of the trading trading day, not qualitatively consistent with the intuition and research supporting the parametric kernel above. Furthermore, the first principal component explains only approximately 1.5% of the total variance in the minutely IOFI matrix. Importantly, fitting PCA on the full sample introduces look-ahead bias that the parametric kernel avoids entirely. We therefore use the exponential decay kernel throughout. Nevertheless, we found the result interesting in its own right.

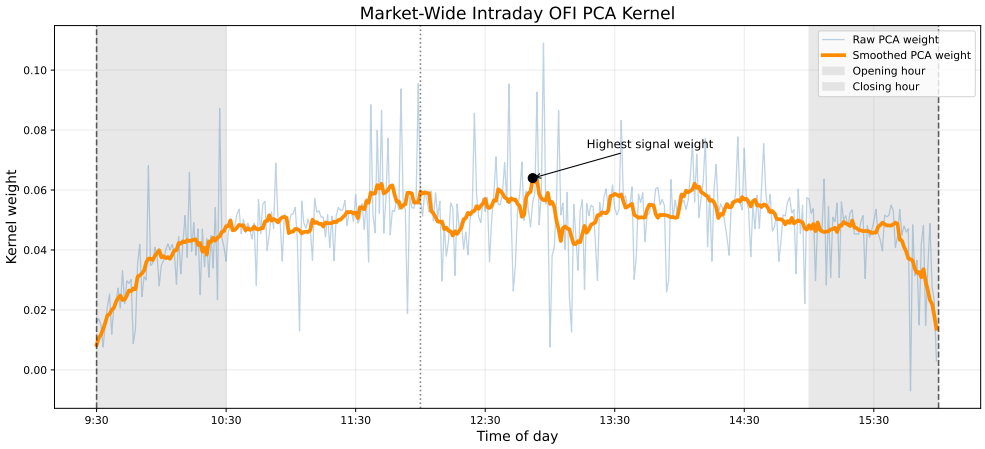


Figure 1: PCA kernel weights across the 390 minutes of the trading day, showing the relative contribution of each minute to the dominant order-flow component.

Finally, with our exponential kernel, we normalize the daily signal using a backward-looking rolling 60-day z-score to place assets on a comparable scale through time, and denote the resulting normalized signal by $\widetilde{\mathrm{ofi}}_{i,t}^{\mathrm{daily}}$.

2.4 Signal and Return Matrices

We assemble the normalized daily OFI signals into a signal matrix

$$P \in \mathbb{R}^{T\times N}, \qquad P_{t,i} = \widetilde{\mathrm{ofi}}_{i,t}^{\mathrm{daily}},$$

where rows correspond to trading dates and columns to assets. In parallel, we construct a residualized return matrix

$$R \in \mathbb{R}^{T\times N}, \qquad R_{t,i} = r_{i,t},$$

where $r_{i,t}$ denotes the residualized open-to-close return of asset i on day t. Raw close-to-close returns are cross-sectionally demeaned and residualized against the first three principal components of the return covariance matrix, estimated on a rolling 252-day window. This removes the market factor and major sector rotations so that all prediction targets are idiosyncratic returns.

2.5 Cross-Asset Payoff Construction

To study one-day-ahead predictability, we lag the signal matrix by one trading day:

$$P^{\mathrm{lag}}_{t,j} := P_{t-1,j}.$$

For each ordered pair (i, j), the realized directed payoff is

$$\mathrm{PNL}_{i,j,t} := R_{t,i}\, P^{\mathrm{lag}}_{t,j},$$

representing the profit from using stock j's lagged OFI to position in stock i. We compute the sample mean $\widehat{\mu}_{ij}$, standard deviation $\widehat{\sigma}_{ij}$, and a Sharpe-like edge score

$$\mathrm{SR}_{ij} := \sqrt{252}\,\frac{\widehat{\mu}_{ij}}{\widehat{\sigma}_{ij}},$$

assembling these into a directed score matrix $A \in \mathbb{R}^{N\times N}$.

3 Preliminary Analysis

Before developing the structured estimation framework in Section 4, we test two approaches to exploiting the score matrix A. Both appear to produce real signal but underperform the regression approach.

3.1 Sparse Directed-Edge Baseline

Our first approach treats A as a directed graph and searches for individual source–target edges with statistically significant predictive content. For each ordered pair (i, j), we compute

$$t_{ij} := \widehat{\mu}_{ij}/(\widehat{\sigma}_{ij}/\sqrt{n_{ij}}),$$

map these to two-sided p-values, and apply Benjamini–Hochberg false discovery rate control at level q (Benjamini and Hochberg, 1995), together with a minimum sample-size requirement `min_count`. On the screened matrix, we alternate between choosing the strongest source columns and strongest target rows until convergence. This is basically a greedy biclustering heuristic for finding the densest predictive submatrix. The resulting signed trading matrix is then applied each day to the lagged signal vector and normalized to unit gross leverage.

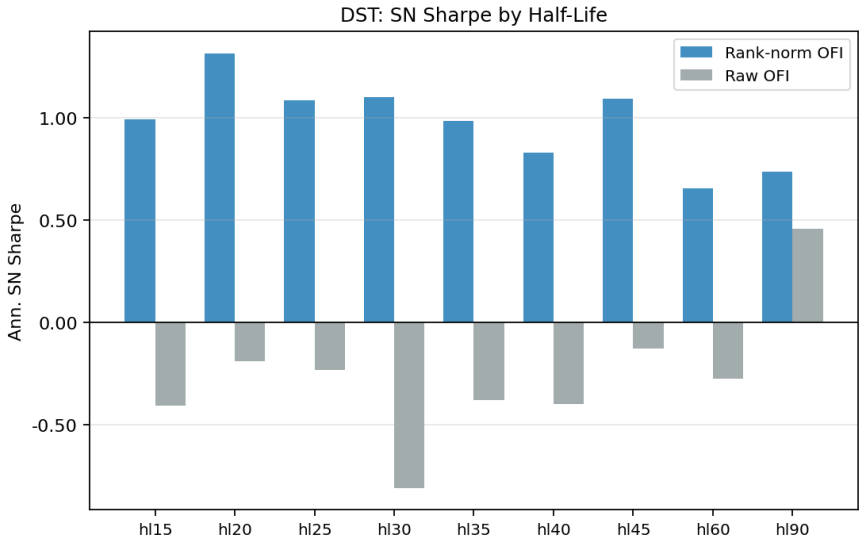


Figure 2: Sharpe ratio by OFI half life parameter (no transaction costs)

With rank-normalized OFI, this method achieves sector-neutral Sharpe ratios of approximately 1.0–1.3 (best: 1.32 at half-life 20 minutes), confirming that cross-asset predictability is real. Importantly, the rank normalization plays a large in making the cross asset OFIs comparable so directly in this matrix form.

3.2 Dhillon Bipartite Spectral Co-Clustering

Rather than treating the predictive graph as a collection of isolated edges, we next ask whether the signal is better understood at the level of *blocks* of related sources and targets. To do this, we use the spectral co-clustering method of Dhillon (2001). This is a natural setting for it: our score matrix is already bipartite, with columns corresponding to source stocks and rows corresponding to target stocks.

For each training window, we first compute the directed edge statistics from lagged OFI signals and future returns. We then form a nonnegative matrix from the absolute edge t-statistics and use it as the input for the co-clustering step. We compute the leading nontrivial singular vectors of the normalized operator and use them to embed rows and columns into a lower-dimensional space. We then run k-means on the stacked row and column embeddings, which gives simultaneous clusters of source assets and target assets.

Once these clusters are fixed, we score each source-cluster $\rightarrow$ target-cluster block by the count-weighted mean of the signed edge Sharpe scores inside that block, and trade the strongest one. Hence, this essentially just aggregates the lagged signals from the source cluster into a single scalar factor, maps that factor onto the target cluster through cluster-level loadings, and then evaluate the resulting positions in a walk-forward backtest, re-clustering about monthly.

In practice, this method only works well once we restrict attention to within-sector edges. Without that restriction, the number of tested source-target pairs is too large. Imposing the sector mask makes the graph more informative and gives stronger results.

The method has some weaknesses. It is not especially stable across re-estimation windows. The chosen block changes fairly often, which leads to high turnover. Also, the procedure is trying to learn both the grouping structure and the predictive relationship from the same noisy sample. In that sense, the within-sector restriction is already doing part of the conceptual work. This is part of what motivates enforcing structure via restrictions on a factor matrix via the regression below.

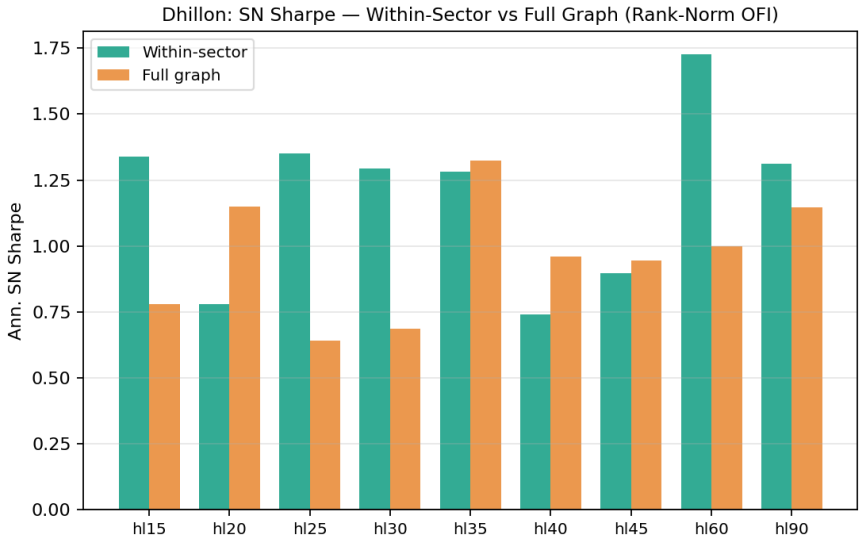


Figure 3: Dhillon co-clustering Sharpe ratio by OFI half-life parameter (no transaction costs). Within-sector masking and rank normalization are both important.

3.3 Summary and Motivation

Table 1 compares the two preliminary graph-based methods.

Table 1: Comparison of the two preliminary graph-based methods. Both models use rank-normalized OFI and are evaluated on next-day residualized returns.

Method	Best model	SN Sharpe	Gross Sharpe	Mean IC
Dhillon co-clustering (within-sector)	`dh_hl60_rn_k2_ws`	+1.73	+0.38	+0.014
Directed source–target	`dst_hl20_rn`	+1.32	+0.86	+0.020

The main pattern is fairly clear. Moving from independent edge selection to a block-based view helps a lot on the sector-neutral metric: the Dhillon approach outperforms the directed source–target baseline by a meaningful margin in SN Sharpe. At the same time, the DST method still has some appeal—its gross Sharpe and mean IC are both a bit stronger, which suggests that there is real signal in the edge-by-edge construction even if it is too noisy and unstable on its own.

What this comparison seems to show is that some kind of block structure is genuinely there. The sparse edge-by-edge picture isn't too strong, but once related sources and targets are grouped together, the signal becomes much cleaner. At the same time, Dhillon only really works once we restrict attention to within-sector edges, so the grouping structure is not being learned from scratch in a completely convincing way. Again, this is why of methods that capture a predictive linear relationship, a regression approach as below is defensible.

4 Main Analysis - Regression Methods

4.1 Cross-Impact Matrix Estimation

Our main object is a cross-impact matrix $W \in \mathbb{R}^{N\times N}$ that maps lagged OFI signals to next-day returns. Specifically, we model

$$r_{t+1} \approx W Z_t,$$

where $r_{t+1} \in \mathbb{R}^N$ is the vector of next-day residual returns and $Z_t \in \mathbb{R}^N$ is the lagged daily OFI signal vector. The entry W_{ij} captures the predictive effect of asset j's OFI on asset i's next-day return. To do this over more than just one day, we form the matrix R, Z made up of columns of r_{t+1} and Z_t, using a rolling update window. Hence, we are essentially looking for the matrix $W \in \mathbb{R}^{N\times N}$ such that

$$W = \arg\min_W \|R - WZ\|_F$$

But, to sparsify, we estimate W using two regularized regression approaches:

Ridge regression: We minimize the squared prediction error subject to an ℓ_2 (Frobenius) penalty:

$$\hat{W}_{\text{ridge}} = \arg\min_W \sum_{t=1}^{T-1} \|r_{t+1} - WZ_t\|_2^2 + \lambda\|W\|_F^2.$$

This admits a closed-form solution $\hat{W}_{\text{ridge}} = \mathbf{R}\,\mathbf{Z}^\top(\mathbf{Z}\,\mathbf{Z}^\top + \lambda I)^{-1}$ and shrinks all entries of W proportionally toward zero. Hoerl and Kennard (1970)

Lasso regression: To induce sparsity, we minimize the same objective with an element-wise ℓ_1 penalty:

$$\hat{W}_{\text{lasso}} = \arg\min_W \sum_{t=1}^{T-1} \|r_{t+1} - WZ_t\|_2^2 + \lambda\|W\|_1.$$

This forces many entries W_{ij} to exactly zero, producing a sparse network of cross-asset relationships and acting as automatic feature selection.

In both cases, the regularization parameter λ is selected from a grid $\{1, 10, 50, 250, 1000\}$ by maximizing in-sample information coefficient (IC). We find that Lasso regression and ridge regression perform similarly. OLS performs similarly as well. Because Lasso regression is computationally expensive to calculate, we test our results for the most part with Ridge regression. We also consider a low rank approximation of our results which should recover some of the stronger sparsity Lasso regression gives. Tibshirani (1996)

4.2 Structural Constraints on W

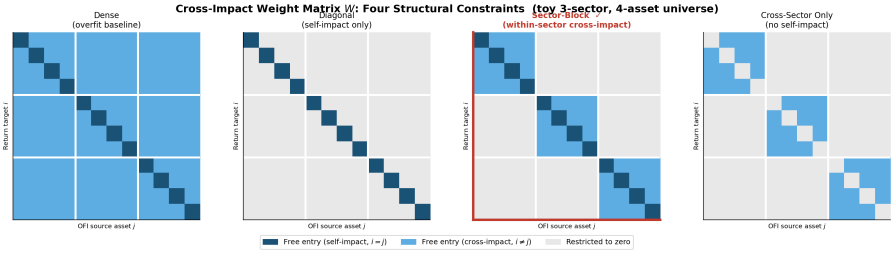
With $N \approx 60$ assets, the full matrix W has approximately $3,600$ parameters, which exceed the number of training observations (≈ 750 days). While we had limited assets with ten levels of limit order book data, we did want our pipeline to function well with far more assets as well, (and even found it worked nearly as well with a few hundred assets with only level 1 OFI information in addition to the ≈ 70 we had with ten levels). Imposing some structure on W then makes sense and fits with our previous findings. We consider four structural constraints, illustrated below:

1. **Dense:** W_{ij} unrestricted for all i, j. This serves as an overfitting baseline.

2. **Diagonal:** $W_{ij} = 0$ for $i \neq j$. Only self-impact (own OFI predicts own return).

3. **Sector-block:** W_{ij} is unrestricted if assets i and j belong to the same GICS sector, and zero otherwise. Cross-impact is permitted within sectors but not across.

4. **Cross-sector only:** W_{ij} is unrestricted only if i and j are in the same sector *and* $i \neq j$. Self-impact is excluded.



In general, we find that the sector block models (based on GICS) are much stronger and more stable than the other variants. As such, the results in the following section will be based on said models.

4.3 Low-Rank Denoising via Gavish–Donoho Thresholding

Even after imposing sector-block structure, the estimated W may retain substantial noise. We apply a post-estimation low-rank filter based on the optimal hard singular value threshold of Gavish and Donoho (2014).

Let $W = U\Sigma V^\top$ be the singular value decomposition of the estimated cross-impact matrix, with singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_N$. The Gavish–Donoho threshold provides an optimal cutoff τ^* below which singular values are consistent with pure noise under a spiked random matrix model. We retain only the singular values exceeding this threshold and form the low-rank approximation

$$W_r = U_r \Sigma_r V_r^\top,$$

where r is the number of retained singular modes. In practice, this retains approximately 10 singular values out of roughly 60 total assets, hinting further that the cross-impact structure low-rank.

4.4 Walk-Forward Evaluation

All models are evaluated in a strictly out-of-sample walk-forward framework. We use a rolling training window of 750 trading days ($\approx$ 3 years) and refit the model every 21 trading days ($\approx$ 1 month). At each refit, we re-estimate W using the most recent 750-day window, apply the Gavish–Donoho filter, and generate predictions for the subsequent 21-day test period.

Signal quality is assessed with a daily cross-sectional long–short rule: on each day, we rank all assets by predicted residual return, go long the top 10%, and short the bottom 10%. We report three main metrics:

- **Spread return:** the raw daily return of the long–short portfolio.
- **Annualized Sharpe ratio:** $\hat{\mu}/\hat{\sigma} \times \sqrt{252}$, computed from the daily spread return series.
- **Information coefficient (IC):** the daily cross-sectional Spearman rank correlation between predicted and realized returns, averaged over time.

We report both the raw spread Sharpe and the sector-neutral (SN) Sharpe. The latter is computed by running the top-10%/bottom-10% long–short independently within each GICS sector and then averaging the resulting sector-level spreads.

4.5 Main Results

Table 2 reports the five best model configurations ranked by annualized sector-neutral Sharpe ratio, computed gross of transaction costs. All models use the sector-block structure with Gavish–Donoho denoising. The final two columns report the breakeven round-trip cost (in basis points) at which net sector-neutral Sharpe reaches zero, and the net sector-neutral Sharpe after a 10 bps round-trip cost assumption.

Table 2: Best model configurations by annualized gross sector-neutral (SN) Sharpe ratio. Breakeven RT is the round-trip transaction cost at which net SN Sharpe equals zero.

Model	SN Sharpe	Spread Sharpe	Mean IC	Breakeven RT	Net SN (1 bp)	Net SN (3 bps)
`ridge_sb_gd_hl30_rn`	+1.68	+0.56	+0.011	13.8 bps	+1.56	+1.32
`ridge_sb_gd_hl40`	+1.67	+0.52	+0.015	13.7 bps	+1.54	+1.30
`ridge_sb_gd_hl90`	+1.59	+0.21	+0.016	12.6 bps	+1.46	+1.21
`ridge_sb_gd_hl15`	+1.54	+0.34	+0.014	12.3 bps	+1.41	+1.16
`ridge_sb_gd_hl60`	+1.51	+0.71	+0.016	12.1 bps	+1.39	+1.14
`ridge_sb` (no GD)	+1.80	+0.59	+0.016	8.0 bps	+1.57	+1.12

The no-GD baseline (`ridge_sb`) achieves the highest gross SN Sharpe of +1.80, but its breakeven of only 8 bps makes it unprofitable at higher realistic transaction costs. Adding the Gavish–Donoho filter reduces gross performance modestly but raises the breakeven to 12–14 bps, yielding positive net Sharpe at 10 bps across all top models. All subsequent analysis focuses on GD models.

4.6 Signal Conclusions

We draw several conclusions regarding signal construction from the full sweep of experiments.

First, the exponential decay half-life matters but the relationship is non-monotone. Half-lives in the range of 30–60 minutes perform consistently well, concentrating weight on the final hour of the trading day. Performance degrades at very short half-lives ($\leq$ 15 min, effectively close-only OFI) and at very long half-lives ($\geq$ 90 min, approximating a uniform intraday average).

Second, rank-normalizing the OFI signal cross-sectionally on each day improves performance for short half-lives but is not universally beneficial. For `hl=30`, rank normalization raises the sector-neutral Sharpe from +1.43 to +1.68; for `hl=90`, it reduces it from +1.59 to +0.96. The benefit appears when there are meaningful intraday outliers that the decay kernel has not yet suppressed; at longer half-lives, averaging already moderates these.

Third, multi-scale signal combinations (combining half-lives of 30 and 90 minutes as a $2N$-feature model) add no value over the single best scale (`ridge_sb_gd_ms3090`: SN Sharpe = +1.21). The signals at different half-lives are highly collinear and offer no complementary information once regularization is applied. Similarly, multi-lag specifications that append the previous day's OFI do not consistently improve performance, confirming that order flow imbalance does not persist meaningfully from one session to the next.

4.7 Model Structure Conclusions

The choice of structural constraint on W is the single most important modeling decision.

Dense models fail. Estimating an unrestricted $W \in \mathbb{R}^{N \times N}$ with $N \approx 63$ assets yields roughly 4,000 free parameters against only 750 training observations, leading to severe overfitting. Dense ridge produces a sector-neutral Sharpe of approximately -0.80, worse than no model at all.

Sector-block structure is essential. Restricting cross-impact to within-sector pairs collapses the effective parameter count to $\sum_k n_k^2 \approx 500$ and yields the strongest results across all metrics in which we take into account cross impact. This implies that OFI-based cross-impact relationships are predominantly sector-localized: the order flow of one technology stock is informative about the return of another technology stock, but not of a utility stock.

Self-impact contributes heavily. Zeroing the diagonal of W (excluding own-OFI from the prediction) reduces performance, indicating that self-impact is not redundant even after cross-asset effects are captured. Furthermore, as above, own asset impact is persistent and strong.

4.7.1 Next-Day vs Overnight Behavior

An interesting pattern emerges when comparing next-day and overnight prediction. For next-day open-to-close returns, the diagonal models — which use only an asset's own OFI — consistently perform best. Cross-asset structure does not appear to add meaningful predictive power at this horizon, and in some cases introduces additional noise.

In contrast, overnight close-to-open prediction shows a different behavior. Here, cross-asset structure — particularly sector-block and cross-sector models — often improves gross performance relative to diagonal models. This suggests that cross-asset order flow information is most relevant at shorter horizons.

One possible interpretation is that cross-asset effects are short-lived. Order flow in one stock may propagate quickly to related stocks within the same sector, affecting overnight gap risk, but this information largely dissipates by the following trading session. By contrast, an asset's own order flow appears to have more persistent predictive content into the next trading day.

This distinction is consistent with Cont et al. (2023), which finds that off-diagonal effects exist but are typically weaker and importantly shorter-lived than self-impact.

4.8 Low-Rank Structure via Gavish–Donoho Denoising

The Gavish–Donoho singular value filter consistently improves out-of-sample performance and transaction-cost robustness. Across rolling windows, the filter retains on average approximately 8–10 singular values from the fitted W matrix (out of a maximum of $N \approx 63$, so roughly 13–16% of the spectral mass). This is consistent with a sparse latent factor structure: a small number of latent factors, perhaps reflecting sector-level supply–demand dynamics or common institutional flows, account for the bulk of the cross-impact signal. The remaining singular values are treated as noise and discarded.

As noted above, Gavish–Donoho denoising reduces turnover and raises the breakeven transaction cost from approximately 8 bps to 12–14 bps, a material improvement for practical deployment.

4.9 Summary of Findings

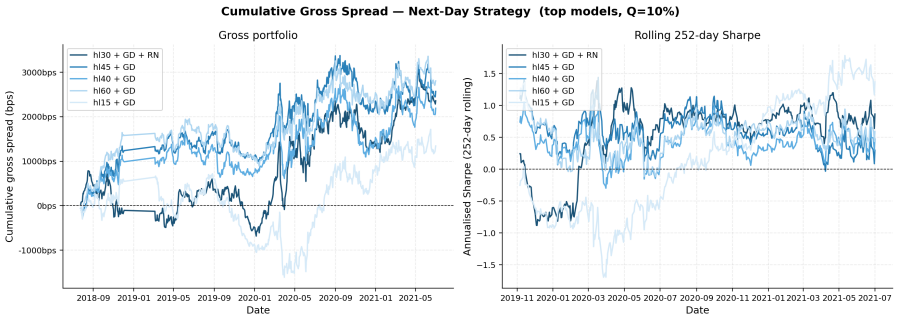


Figure 4: Cumulative PNL & Rolling Sharpe ratio for next-day sector block models

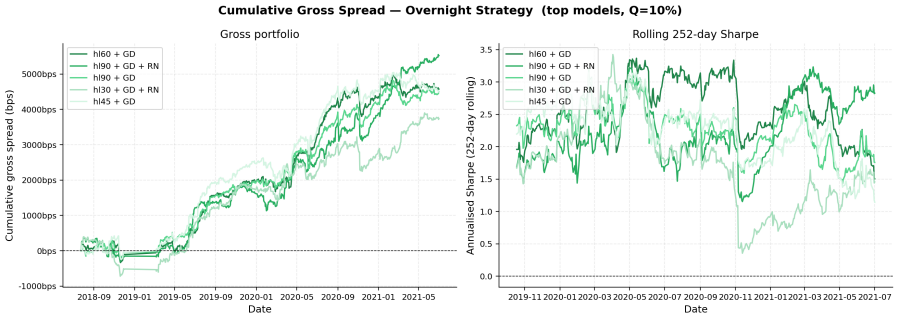


Figure 5: Cumulative PNL & Rolling Sharpe ratio for overnight sector block models

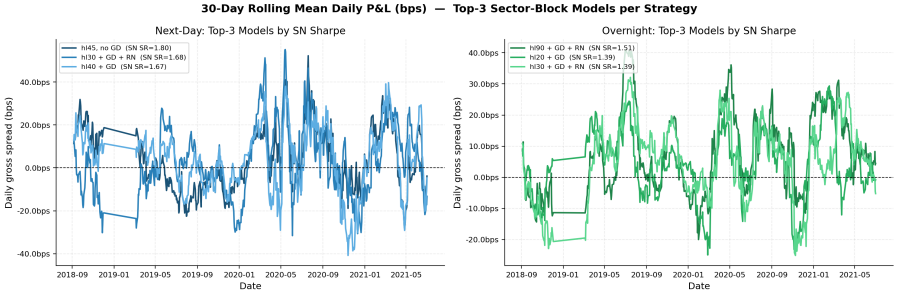


Figure 6: Average Rolling Daily PNL for sector block models

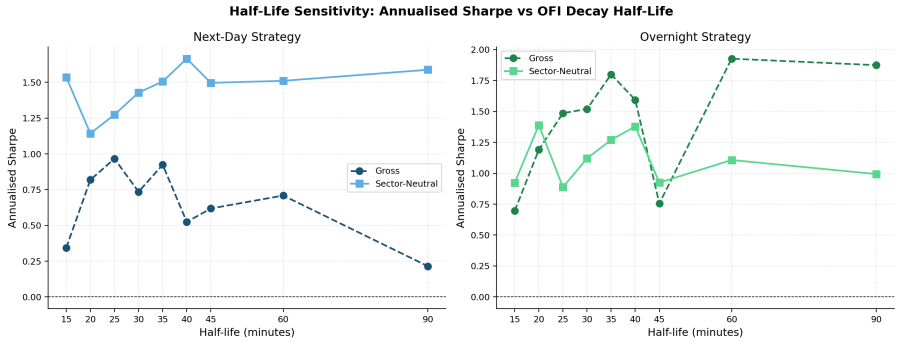


Figure 7: Annualized Sharpe ratio as a function of the OFI exponential decay half-life (minutes), for the next-day (left) and overnight (right) strategies.

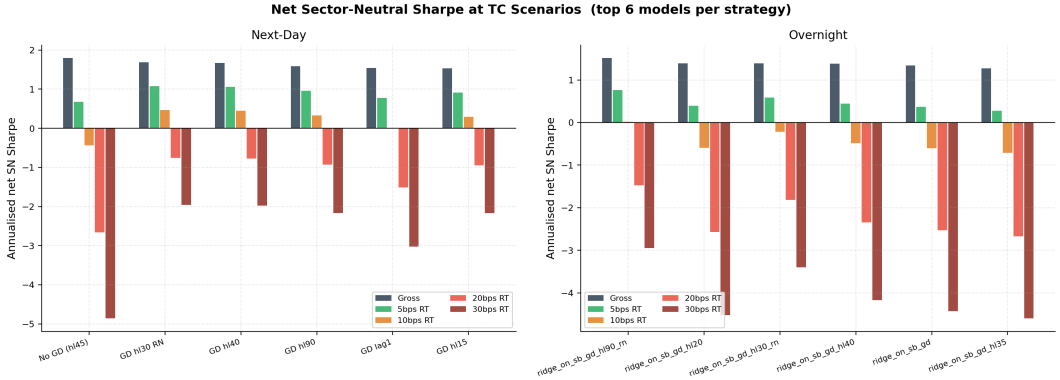


Figure 8: Net sector-neutral Sharpe ratio under round-trip transaction cost scenarios of 0, 10, 20, and 30 bps, for the top models in the next-day (left) and overnight (right) strategies.

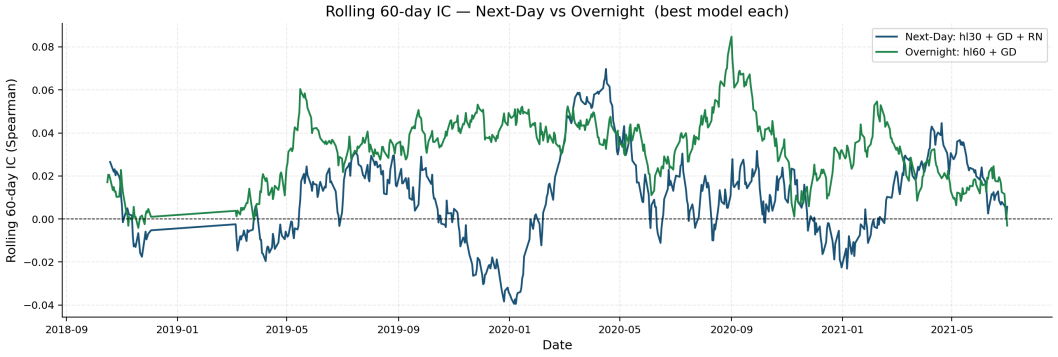


Figure 9: Rolling 60-day mean Spearman IC for the best next-day (`hl30_rn`) and overnight (`hl90_rn`) ridge models over the out-of-sample evaluation period. Persistent positive IC across regimes indicates the OFI signal is not period-specific.

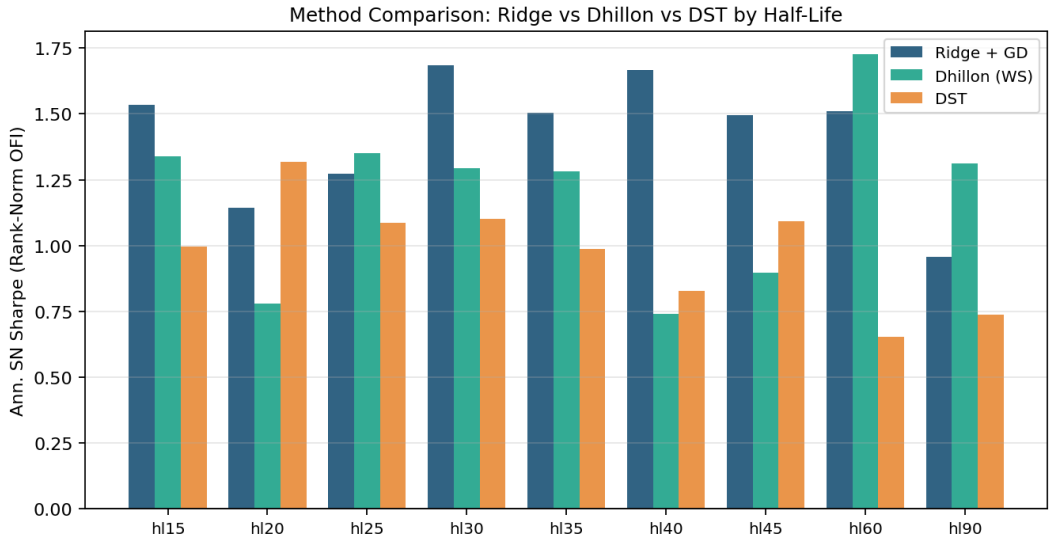


Figure 10: Annualized sector-neutral spread Sharpe ratio by half-life for Ridge + Gavish–Donoho, Dhillon bipartite co-clustering (within-sector), and directed source-target biclustering, all using rank-normalized OFI signals. Ridge dominates at shorter half-lives; all three methods peak in the 20–35 minute range.

To recap, the key takeaways are:

1. **Overnight and Next Day Signal is real.** There exists substantial predicative signal for overnight and next day returns in integrated OFI data, albeit signals with some different characteristics.
2. **Structure over estimator.** The structural constraint on W matters far more than the choice of ridge versus lasso versus OLS.
3. **Cross-impact is sector-local.** Within-sector cross-asset effects drive the signal; cross-sector effects are negligible.
4. **Cross-impact exceeds self-impact in certain cases.** Diagonal models underperform sector-block models overnight, confirming that cross-asset order flow contains materially more predictive information than own-asset order flow alone. However, next day, diagonal models perform somewhat better that sector block.
5. **The signal is low-rank.** Gavish–Donoho thresholding retains roughly 15% of singular values, consistent with a small number of latent factors driving intra-sector cross-impact.

5 Conclusion

We have shown that intraday order flow imbalance contains information that persists to the next trading day and can be used to predict daily stock returns, not only within an asset but across assets. The key to extracting this signal is structural: imposing sector-block constraints on the cross-impact matrix and applying low-rank denoising via singular value thresholding. The best model achieves an annualized sector-neutral Sharpe ratio of approximately 1.6 out of sample.

Several findings stand out. Cross-asset OFI effects are predominantly sector-local and low-rank, consistent with a small number of latent factors driving intra-sector supply–demand dynamics. The choice of structural constraint on the cross-impact matrix matters far more than the choice of estimator. Rank normalization

of the raw OFI signal is critical for robustness. And the predictive signal does not persist beyond a single day—multi-lag specifications add no value.

Natural extensions include data-driven clustering to discover source–target groups beyond GICS sectors, multi-level OFI that incorporates depth beyond the best quotes, and nonlinear models that may capture more complex cross-asset interactions.

References

Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300, 1995. doi: 10.1111/j.2517-6161.1995.tb02031.x. URL https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Stefanos Bennett, Mihai Cucuringu, and Gesine Reinert. Lead-lag detection and network clustering for multivariate time series with an application to the us equity market. *Machine Learning*, 111(12):4497–4538, 2022.

M Benzaquen, I Mastromatteo, Z Eisler, and J-P Bouchaud. Dissecting cross-impact on stock markets: an empirical analysis. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(2):023406, feb 2017. doi: 10.1088/1742-5468/aa53f7. URL https://doi.org/10.1088/1742-5468/aa53f7.

Francesco Capponi and Rama Cont. Multi-asset market impact and order flow commonality. *SSRN Electronic Journal*, 10 2020. URL https://ssrn.com/abstract=3706390.

Álvaro Cartea, Mihai Cucuringu, and Zijun Jin. Detecting lead-lag relationships in stock returns and portfolio strategies, 2023. SSRN working paper.

Tarun Chordia and Avanidhar Subrahmanyam. Order imbalance and individual stock returns: Theory and evidence. *Journal of Financial Economics*, 72(3):485–518, June 2004. doi: 10.1016/S0304-405X(03)00175-2. URL https://doi.org/10.1016/S0304-405X(03)00175-2.

R. Cont, A. Kukanov, and S. Stoikov. The price impact of order book events. *Journal of Financial Econometrics*, 12(1):47–88, June 2013. ISSN 1479-8417. doi: 10.1093/jjfinec/nbt003. URL http://dx.doi.org/10.1093/jjfinec/nbt003.

Rama Cont, Mihai Cucuringu, and Chao Zhang. Cross-impact of order flow imbalance in equity markets. *Quantitative Finance*, 23(10):1373–1393, October 2023. doi: 10.1080/14697688.2023.2236159. URL https://doi.org/10.1080/14697688.2023.2236159.

Inderjit S. Dhillon. Co-clustering documents and words using bipartite spectral graph partitioning. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 269–274. ACM, 2001. doi: 10.1145/502512.502550. URL https://doi.org/10.1145/502512.502550.

Matan Gavish and David L. Donoho. The optimal hard threshold for singular values is 4/sqrt(3), 2014. URL https://arxiv.org/abs/1305.5870.

Joel Hasbrouck. Measuring the information content of stock trades. *The Journal of Finance*, 46(1):179–207, March 1991. doi: 10.1111/j.1540-6261.1991.tb03749.x. URL https://doi.org/10.1111/j.1540-6261.1991.tb03749.x.

Yixuan He, Gesine Reinert, and Mihai Cucuringu. Digrac: Digraph clustering based on flow imbalance. 2022. URL https://arxiv.org/abs/2106.05194.

Arthur E. Hoerl and Robert W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970. doi: 10.1080/00401706.1970.10488634.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, 58(1):267–288, 1996. doi: 10.1111/j.2517-6161.1996.tb02080.x.